

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1 Metode Pengumpulan Data**

Metode pengumpulan data adalah cara-cara yang digunakan untuk mengumpulkan informasi yang diperlukan dalam suatu penelitian atau studi. Pemilihan metode yang tepat sangat penting untuk memastikan bahwa data yang dikumpulkan akurat, relevan, dan dapat diandalkan. Pengumpulan data yang dilakukan dalam penelitian ini adalah sebagai berikut:

##### **3.1.1 Tahapan Pengamatan (*Observasi*)**

Pengamatan data merupakan metode pengumpulan informasi yang dilakukan dengan mengamati dan mencatat data secara langsung di lapangan. Dalam penelitian ini, pengamatan dilakukan di SMA Negeri 1 Pagelaran yang terletak di Jl.Raya Gumukrejo, Kec. Pagelaran, Kabupaten Pringsewu, Lampung. Beberapa atribut yang diamati meliputi Data Akademik, yang memberikan wawasan mengenai strategi mutu sekolah. Selain itu, jenis kelamin peserta dicatat untuk memahami proporsi antara laki-laki dan perempuan dalam penerimaan, sementara pilihan jurusan yang diambil mencerminkan minat dan kecenderungan akademis mereka. Data yang dikumpulkan melalui pengamatan ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang karakteristik dan nilai akademik di SMA Negeri 1 Pagelaran.

##### **3.1.2 Tahapan Wawancara**

Tahap wawancara adalah metode pengumpulan data yang melibatkan tanya jawab langsung dengan narasumber mengenai topik tertentu. Dalam konteks ini, Wawancara dilakukan dengan Waka Kurikulum untuk mengeksplorasi strategi Meningkatkan Mutu Sekolah yang diterapkan oleh SMA Negeri 1 Pagelaran. Pertanyaan yang diajukan mencakup analisis data akademik, seperti nilai akademik,

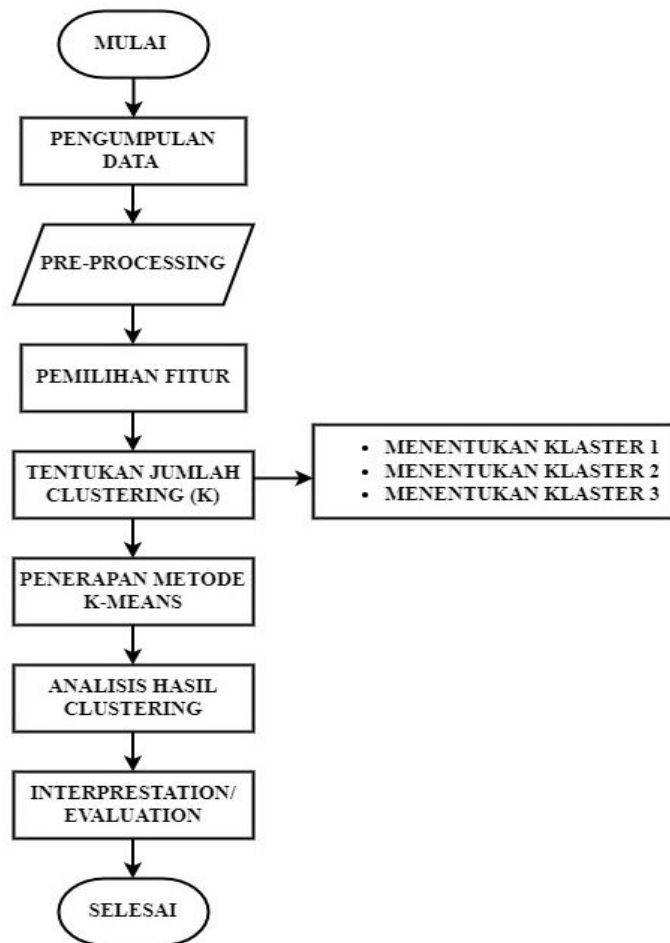
nilai non akademik, status alumni dan jenis kelamin. guna memahami efektivitas Meningkatkan Mutu Sekolah.

### **3.1.3 Tahapan Studi Pustaka**

Tinjauan pustaka merujuk pada kajian *teoritis* dan *referensi* lain mengenai nilai, budaya, dan norma yang berkembang dalam konteks sosial penelitian tersebut. Selain itu, studi pustaka sangat penting dalam penelitian Karena penelitian tidak dapat dipisahkan dari *literature ilmiah*. Hal ini mencakup jurnal-jurnal yang berkaitan dengan *Data Mining* menggunakan Algoritma *K-Means Clustering*.

### **3.2 Tahapan Penelitian**

Metode penelitian yang diterapkan dalam studi ini dijelaskan dalam alur penelitian. Alur penelitian menggambarkan langkah-langkah yang dilakukan dari awal hingga akhir. Beberapa tahapan yang akan dilakukan dalam penelitian ini meliputi pengelompokan Jurusan, Jenis Kelamin, Nilai akademik, Nilai non akademik dan Status Alumni. Dengan menggunakan *K-Means Clustering* yang dapat dilihat sebagai berikut :



Gambar 3. 1 Tahap Penelitian

### 3.2.1 Tahap Mengumpulkan Data Akademik

Tahapan ini peneliti mengumpulkan data akademik di SMA Negeri 1 Pagelaran. Data yang dikumpulkan mencakup informasi dari data akademik untuk Tahun Ajaran 2020/2021, 2021/2022 dan 2022/2023. *Dataset* tersebut terdiri dari 982 data siswa, dengan atribut meliputi jenis kelamin, nama, jurusan, nilai pengetahuan, keterangan, dan status alumni. Pengumpulan data dilakukan pada tanggal 9 Desember 2024. *Dataset* bisa dilihat pada Tabel 3.1.

Tabel 3. 1Data Mentah (981 baris)

| Nama                       | Jenis Kelamin | Jurusan | Nilai Akd | Nilai Non/Akd | Status  |
|----------------------------|---------------|---------|-----------|---------------|---------|
| Alesya Al Fadilla          | perempuan     | ipa     | 77.20     | 77.20         | Bekerja |
| Abel Tri Yunita            | perempuan     | ipa     | 77.80     | 78.07         | Bekerja |
| Adelia Dwi Puspita Ningrum | perempuan     | ipa     | 80.27     | 80.13         | Bekerja |
| Adilia Dwi Wiratma         | perempuan     | ipa     | 77.87     | 77.60         | Bekerja |
| Adinda Intania Iskandar    | laki          | ipa     | 76.73     | 77.20         | Bekerja |
| Aditya Dwi Maulana         | perempuan     | ipa     | 83.47     | 83.07         | Kuliah  |
| Afifah Taskia              | laki          | ipa     | 81.47     | 81.07         | Bekerja |
| Ages Seprina               | perempuan     | ipa     | 81.00     | 80.67         | Kuliah  |
| Ajeng Fibri Prastika       | laki          | ips     | 78.33     | 77.73         | Bekerja |
|                            |               |         |           |               |         |

### 3.2.2 Pre-processing

*Pre-processing* adalah proses persiapan data sebelum analisis atau pemodelan, yang meliputi seleksi, pembersihan, dan perubahan data. **Seleksi data** dilakukan untuk memilih data yang relevan agar analisis lebih efisien. **Pembersihan data** bertujuan mengatasi masalah seperti duplikasi, nilai yang hilang, atau kesalahan format agar kualitas data lebih baik. **Perubahan data** mencakup transformasi seperti normalisasi, *encoding*, dan reduksi dimensi agar data lebih sesuai untuk diproses. Dengan *preprocessing* yang baik, data menjadi lebih bersih, terstruktur, dan siap digunakan untuk analisis atau model *machine learning* sebagai berikut :

#### a. Seleksi Data (*Data Selection*)

Data yang didapat dari objek penelitian data berupa data Akademik di Tahun Ajaran 2020/2021, 2021/2022, 2022/2023, akan dilakukan seleksi data atau atribut yang relevan untuk penelitian. Sehingga didapatkan atribut-atribut yang akan diseleksi untuk penelitian ini yaitu atribut Jenis Kelamin, Jurusan, Nilai akademik, Nama, Nilai non akademik dan Status Alumni. Setelah menentukan seleksi atribut yang akan digunakan maka data akademik akan dikelompokkan berdasarkan kolom Jenis Kelamin, kolom Jurusan, kolom Nilai akademik, Kolom Nilai non akademik dan kolom Status Alumni dimana data akademik yang dilakukan data *selection*, pada tahun ajaran 2020/2021,2021/2022,2022/2023. Contoh Hasil Seleksi Data :

Tabel 3. 2 Contoh Hasil Seleksi Data

| No | Jurusan | Jenis Kelamin | Nilai Akd | Nilai Non-Akd | Status  |
|----|---------|---------------|-----------|---------------|---------|
| 1. | Ipa     | Laki-Laki     | 77.33     | 77.20         | Bekerja |
| 2. | Ipa     | Perempuan     | 77.80     | 78.07         | Bekerja |
| 3. | Ipa     | Perempuan     | 80.27     | 80.13         | Bekerja |
| 4. | Ipa     | Perempuan     | 77.87     | 77.60         | Bekerja |
| 5. | Ipa     | Laki-Laki     | 76.73     | 77.20         | Kuliah  |
| 6  | Ipa     | Perempuan     | 83.47     | 83.07         | Bekerja |
| 7  | Ipa     | Laki-Laki     | 81.47     | 81.07         | Kuliah  |
| 8. | Ipa     | Perempuan     | 81.00     | 80.67         | Bekerja |
| 9. | Ips     | Laki-Laki     | 78.33     | 77.73         | Kuliah  |

b. Pembersihan Data (*Cleansing Data*)

Proses pembersihan data dilakukan untuk mengatasi duplikasi dan *missing value* dalam *dataset*. Ditemukan sebanyak 76 data duplikat, yang kemudian dihapus menggunakan fitur "*Remove Duplicates*", mengurangi jumlah *dataset* dari 981 menjadi 905. Selain itu, terdapat satu *missing value* pada atribut Nilai Non Akademik, yang bertipe data real. *Missing value* ini menunjukkan adanya data yang tidak lengkap, sehingga perlu ditangani agar analisis dapat dilakukan secara optimal. Untuk mengatasi masalah ini, digunakan metode "*Replace Missing Value*" dengan mengganti nilai kosong pada atribut tersebut dengan nilai tertentu. Langkah-langkah ini bertujuan untuk meningkatkan kualitas *dataset* agar lebih akurat dan siap digunakan dalam proses analisis lebih lanjut. Contoh Hasil "*Remove Duplicates*" dan "*Replace Missing Value*" bisa dilihat pada Tabel beriku

Tabel 3. 3 Contoh Sebelum *Missing Value*

| Name               | Type    | Missing | Statistic   | Filter       |
|--------------------|---------|---------|-------------|--------------|
| Jenis Kelamin      | Nominal | 0       | P(2)        | P(623)       |
| Jurusan            | Nominal | 0       | Ips(81)     | Ipa(593)     |
| Nilai Non Akademik | Real    | 0       | 72          | 88           |
| Nilai Akademik     | Real    | 1       | 72.60       | 88           |
| Status Alumni      | Nominal | 0       | Kuliah(475) | Bekerja(506) |

Tabel 3. 4 Contoh Hasil *Replace Missing Value*

| Name | Type | Missing | Statistic | Filter |
|------|------|---------|-----------|--------|
|------|------|---------|-----------|--------|

|                    |         |   |             |              |
|--------------------|---------|---|-------------|--------------|
| Jenis Kelamin      | Nominal | 0 | P(2)        | P(623)       |
| Jurusan            | Nominal | 0 | Ips(81)     | Ipa(593)     |
| Nilai Non Akademik | Real    | 0 | 72          | 88           |
| Nilai Akademik     | Real    | 0 | 72.60       | 88           |
| Status Alumni      | Nominal | 0 | Kuliah(475) | Bekerja(506) |

Tabel 3. 5 Contoh Statistik Data Sebelum di *Remove Duplikat*

| Kolom              | Tipe        | Statistik      |               |                               |
|--------------------|-------------|----------------|---------------|-------------------------------|
|                    |             | Paling Sedikit | Paling Banyak | Nilai                         |
| Jenis Kelamin      | Polynominal | L (356)        | P (625)       | L(356), P(625)                |
| Jurusan            | Polynominal | IPS(388)       | IPA(593)      | IPS(388),<br>IPA(593)         |
| Nilai Akademik     | Real        | 72             | 88            | Avg. 79.222                   |
| Nilai Non Akademik | Real        | 72.020         | 88            | Avg. 79.151                   |
| Status Alumni      | Polynominal | Kuliah (475)   | Bekerja(506)  | Kuliah (475),<br>Bekerja(506) |

Tabel 3. 6 Contoh Hasil *Remove Duplicates*

| Kolom              | Tipe        | Statistik      |               |                               |
|--------------------|-------------|----------------|---------------|-------------------------------|
|                    |             | Paling Sedikit | Paling Banyak | Nilai                         |
| Jenis Kelamin      | Polynominal | L (338)        | P (558)       | L(338), P(558)                |
| Jurusan            | Polynominal | IPS(378)       | IPA(518)      | IPS(388),<br>IPA(593)         |
| Nilai Akademik     | Real        | 72             | 88            | Avg. 79.059                   |
| Nilai Non Akademik | Real        | 72.020         | 88            | Avg. 78.978                   |
| Status Alumni      | Polynominal | Kuliah (431)   | Bekerja(465)  | Kuliah (475),<br>Bekerja(506) |

c. Perubahan Data (*Transformation Data*)

*Data transformation* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data Mining*. Pada langkah *transformation* dilakukan untuk data bertipe *polynomial* menjadi numerik agar data dapat diolah berdasarkan jarak. Untuk mengubah data bertipe *polynomial* menjadi numerik menggunakan *operator Nominal to Numerical*. Oleh sebab itu dataset penerimaan peserta didik baru akan ditransformasikan ke dalam bentuk angka pada beberapa atribut yang terdapat huruf, dan pilihan 3 atribut yang digunakan dalam proses *clustering* yaitu atribut jenis kelamin, jurusan dan status alumni.

Tabel 3. 7 Contoh Data Sebelum *Transformation Data*

| No | Jurusan | Jenis Kelamin | Nilai Akd | Nilai Non-Akd | Status Alumni |
|----|---------|---------------|-----------|---------------|---------------|
| 1. | Ipa     | Laki-Laki     | 77.33     | 77.20         | Bekerja       |
| 2. | Ipa     | Perempuan     | 77.80     | 78.07         | Bekerja       |
| 3. | Ipa     | Perempuan     | 80.27     | 80.13         | Bekerja       |
| 4. | Ipa     | Perempuan     | 77.87     | 77.60         | Bekerja       |
| 5. | Ipa     | Laki-Laki     | 76.73     | 77.20         | Kuliah        |
| 6  | Ipa     | Perempuan     | 83.47     | 83.07         | Bekerja       |
| 7  | Ipa     | Laki-Laki     | 81.47     | 81.07         | Kuliah        |
| 8. | Ipa     | Perempuan     | 81.00     | 80.67         | Bekerja       |
| 9. | Ips     | Laki-Laki     | 78.33     | 77.73         | Kuliah        |

Dalam proses pengolahan *dataset*, dilakukan transformasi data untuk mengubah atribut kategorikal (Jenis Kelamin, Jurusan, dan Status Alumni) menjadi format numerik agar lebih mudah dianalisis. Algoritma *Data Mining*, seperti *K-Means Clustering*, lebih efektif dalam mengolah data numerik dibandingkan teks. Transformasi ini menggunakan *Label Encoding*, di mana setiap kategori diganti dengan angka, misalnya "Laki-laki" = 1 dan "Perempuan" = 0 pada atribut Jenis Kelamin. Hal serupa diterapkan pada Jurusan dan Status Alumni dengan kode

angka unik. Konversi ini meningkatkan efisiensi analisis, mempercepat pemodelan, serta memastikan *dataset* dapat digunakan secara optimal dalam teknik *Clustering* dan pengolahan data lebih lanjut. Bisa dilihat pada Tabel dibawah

Tabel 3.8 Contoh Data Sesudah Transformation Data

| No | Jurusan | Jenis Kelamin | Nilai Akd | Nilai Non-Akd | Status Alumni |
|----|---------|---------------|-----------|---------------|---------------|
| 1. | 0       | 1             | 77.33     | 77.20         | 0             |
| 2. | 0       | 0             | 77.80     | 78.07         | 0             |
| 3. | 0       | 0             | 80.27     | 80.13         | 0             |
| 4. | 0       | 0             | 77.87     | 77.60         | 0             |
| 5. | 0       | 1             | 76.73     | 77.20         | 1             |
| 6. | 0       | 0             | 83.47     | 83.07         | 0             |
| 7. | 0       | 1             | 81.47     | 81.07         | 1             |
| 8. | 0       | 0             | 81.00     | 80.67         | 0             |
| 9. | 1       | 1             | 78.33     | 77.73         | 1             |

#### d. Pemilihan Fitur

Pemilihan fitur adalah proses memilih atribut yang paling relevan untuk meningkatkan efisiensi dan akurasi model. Misalnya, dari *dataset* dengan atribut Jenis Kelamin, Jurusan, Nilai Akademik, Nilai Non-Akademik, dan Status Alumni. Sementara atribut yang kurang relevan dapat dihapus. Dengan cara ini, model menjadi lebih sederhana, cepat, dan akurat dalam melakukan prediksi. Bisa dilihat pada Tabel dibawah :

Tabel 3.9 Contoh Hasil Pemilihan Fitur

| No | Jurusan | Jenis Kelamin | Nilai Akd | Nilai Non Akd | Status Alumni |
|----|---------|---------------|-----------|---------------|---------------|
| 1. | Ipa     | Laki-Laki     | 77.33     | 77.20         | Bekerja       |
| 2. | Ipa     | Perempuan     | 77.80     | 78.07         | Bekerja       |
| 3. | Ipa     | Perempuan     | 80.27     | 80.13         | Bekerja       |
| 4. | Ipa     | Perempuan     | 77.87     | 77.60         | Bekerja       |



### 3.2.3 K-Means Clustering

Tahap *Data Mining* dilakukan pengolahan data akademik menggunakan teknik *K-Means Clustering* di *RapidMiner*. Metode ini mengelompokkan data berdasarkan kemiripan nilai untuk mengidentifikasi pola distribusi prestasi siswa. *K-Means* bekerja dengan menentukan pusat *cluster* (*centroid*) secara acak, mengelompokkan data berdasarkan jarak terdekat, lalu memperbarui *centroid* hingga stabil. Keunggulannya adalah efisiensi dalam mengelola data besar dan mengungkap pola tersembunyi, yang mendukung analisis akademik lebih lanjut. Contoh Hasil *K-Means Clustering*. Contoh Hasil *K-Means Clustering* bisa dilihat pada gambar berikut:

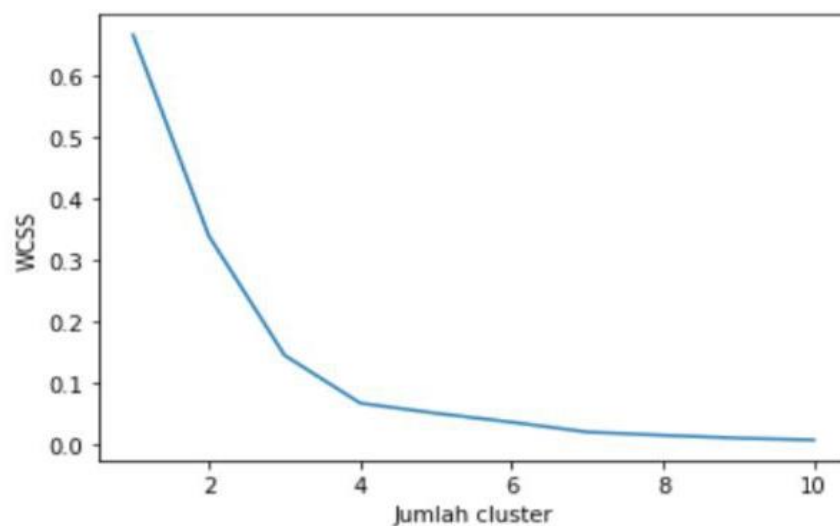
Tabel 3. 10 Contoh Hasil *K-Means Clustering*

| No | Cluster   | Jenis Kelamin | Jurusan | Status | Nilai Non-Akd | Nilai Akd |
|----|-----------|---------------|---------|--------|---------------|-----------|
| 1. | Cluster_2 | 1             | 0       | 0      | 77.80         | 78.70     |
| 2. | Cluster_2 | 1             | 0       | 0      | 80.27         | 80.13     |
| 3. | Cluster_2 | 1             | 0       | 0      | 77.87         | 77.60     |
| 4. | Cluster_1 | 0             | 0       | 1      | 76.73         | 77.20     |
| 5. | Cluster_0 | 1             | 0       | 0      | 83.47         | 83.70     |
| 6. | Cluster_0 | 0             | 0       | 1      | 81.47         | 81.70     |
| 7. | Cluster_2 | 1             | 1       | 0      | 81            | 80.67     |
| 8. | Cluster_2 | 0             | 0       | 1      | 78.33         | 77.73     |

### 3.2.4 Interpretasi/Evaluasi

Tahap *Interpretation/evaluation* atau tahap pengujian menggunakan *dataset* Akademik yang mendaftar di SMA Negeri 1 Pagelaran, data tersebut nantinya diolah menggunakan *software Rapidminer* agar lebih mudah dipahami. Metode kinerja model *Cluster* menghasilkan grafik yang membentuk siku, di mana lekukan pada grafik tersebut menunjukkan jumlah *Cluster* optimal. Oleh karena itu, dapat

membantu dalam menentukan jumlah *Cluster* yang tepat. Pada proses *clustering*, dihitung jarak terpendek atau minimum antara data dan *centroid*, serta jarak maksimum antara *centroid* menggunakan *Within Cluster Sum of Squares* (WCSS). WCSS merupakan total jarak antara anggota *Cluster* terhadap *centroid*, dan mengukur seberapa baik data dikelompokkan dalam setiap *Cluster*. Untuk menghitung WCSS, dilakukan penjumlahan dari seluruh jarak data dikurangi dengan *centroid Cluster* yang bersangkutan, kemudian dikuadratkan. Contoh Hasil visualisasi bisa dilihat pada gambar berikut :



Gambar 3. 2 Contoh Hasil *Elbow Method*

Grafik di atas adalah hasil dari metode "*Elbow*" yang digunakan untuk menentukan jumlah *Cluster* optimal dalam analisis klastering, seperti *K-Means*. Pada grafik ini, sumbu horizontal (X) menunjukkan jumlah *Cluster* yang diuji, sementara sumbu vertikal (Y) menunjukkan nilai WCSS (*Within-Cluster Sum of Squares*), yang menggambarkan seberapa baik data dalam setiap *Cluster* dikelompokkan. WCSS akan menurun seiring bertambahnya jumlah *Cluster*, tetapi penurunan tersebut akan melambat pada titik tertentu, membentuk sebuah sudut atau "*Elbow*". Titik ini menunjukkan jumlah *Cluster* optimal karena setelahnya penambahan *Cluster* tidak memberikan pengurangan WCSS yang signifikan. Berdasarkan grafik, titik *elbow*

tampaknya berada pada *Cluster* ke-3 atau ke-4, sehingga jumlah *Cluster* optimal adalah 3 atau 4.

### 3.3 Sampel Data

Sampel data akademik di SMA Negeri 1 Pagelaran pada tahun 2020-2023. Contoh data dalam format Excel sebagai berikut:

Tabel 3. 11 Sampel Data

| Nama                       | Jenis Kelamin | Jurusan | Nilai Akd | Nilai Non/Akd | Status  |
|----------------------------|---------------|---------|-----------|---------------|---------|
| Alesya Al Fadilla          | perempuan     | ipa     | 77.20     | 77.20         | Bekerja |
| Abel Tri Yunita            | perempuan     | ipa     | 77.80     | 78.07         | Bekerja |
| Adelia Dwi Puspita Ningrum | perempuan     | ipa     | 80.27     | 80.13         | Bekerja |
| Adilia Dwi Wiratma         | perempuan     | ipa     | 77.87     | 77.60         | Bekerja |
| Adinda Intania Iskandar    | laki          | ipa     | 76.73     | 77.20         | Bekerja |
| Aditya Dwi Maulana         | perempuan     | ipa     | 83.47     | 83.07         | Kuliah  |
| Afifah Taskia              | laki          | ipa     | 81.47     | 81.07         | Bekerja |
| Ages Seprina               | perempuan     | ipa     | 81.00     | 80.67         | Kuliah  |
| Ajeng Fibri Prastika       | laki          | ips     | 78.33     | 77.73         | Bekerja |

